ELSEVIER

Gene 281 (2001) 11–17

GENE
AN INTERNATIONAL JOURNAL ON GENES AND GENOMES

www.elsevier.com/locate/gene

The universal ancestor was a thermophile or a hyperthermophile

Massimo Di Giulio*

International Institute of Genetics and Biophysics, CNR, Via G. Marconi 10, 80125 Naples, Italy

Received 28 September 2001; accepted 5 November 2001
Received by G. Bernardi

Abstract

By exploiting the correlation between the optimal growth temperature of organisms and a thermophily index based on the propensity of amino acids to enter thermophile/hyperthermophile proteins, an analysis is conducted in order to establish whether the last universal common ancestor (LUCA) was a mesophile or a (hyper)thermophile. This objective is reached by using maximum parsimony and maximum likelihood to reconstruct the ancestral sequences of the LUCA for two pairs of sets of paralogous protein sequences by means of the phylogenetic tree topology derived from the small subunit ribosomal RNA, even if this is rooted in all three possible ways. The thermophily index of all the reconstructed ancestral sequences of the LUCA belongs to the set of the thermophile/hyperthermophile sequences, thus supporting the hypotheses that see the LUCA as a thermophile or a hyperthermophile.

Keywords: Paralogous protein; Thermophily index; Ancestral sequence; Last universal common ancestor; Origin of life

1. Introduction

Galtier et al. (1999) introduce a simple idea that makes it possible to differentiate between the mesophilic and thermophilic nature of the last universal common ancestor (LUCA). By exploiting (Galtier et al., 1999) the correlation between the optimal growth temperature of prokaryotes and the G + C content of ribosomal RNAs (Galtier and Lobry, 1997), and estimating the G + C content of the LUCA's ancestral sequence, Galtier et al. (1999) are able to establish whether the LUCA was a mesophile or a thermophile by noting whether this content lies between the mesophilic or thermophilic sequences. Clearly, if we had an equivalent relationship between the optimal growth temperature of organisms and a variable derived from the amino acid composition of proteins, then the idea of Galtier et al. (1999) could be extended to any type of protein. Such a variable does in fact exist (Di Giulio, 2000a) and is based on the differing propensity of amino acids to enter mesophile or thermophile/hyperthermophile proteins (Di Giulio, 2000a). Therefore, in the present paper, I use the strong correlation between the optimal growth temperature of organisms and a thermophily index (Di Giulio, 2000a) and employ methods for the reconstruction of the LUCA's ancestral sequence based on maximum parsimony and maximum likelihood to establish whether the LUCA was a mesophile or a (hyper)thermophile.

2. Materials and methods

The signal recognition particle (SRP) sequences of both the 54 kDa and the subunit α (FtsY) were obtained from the web site http://psyche.uthct.edu/dbs/SRPDB/SRPDB.html. The set of sequences was aligned using the CLUSTALX program (Thompson et al., 1997). The alignment of the sequences of these two paralogous proteins is equal, with the exception of a handful of sites, to the one reported in Gribaldo and Cammarano (1998). All the sites containing at least one gap were eliminated from the alignment, which turned out to be 220 residues long. This elimination was necessary because the algorithm which reconstructs the ancestral sequences by means of maximum likelihood (Zhang and Nei, 1997) cannot deal with gaps.

The alignment of tryptophanyl-tRNA synthetase sequences and tyrosyl-tRNA synthetase sequences is the same as the one reported in Diaz-Lazcoz et al. (1998). After the elimination of sites containing at least one gap, the sequences were 127 amino acids long.

These alignments, like all the other files used in the analysis, are available upon request.

The maximum parsimony criterion used in reconstructing the ancestral sequences was employed using the PAUP 3.1.1

Abbreviations: LUCA, last universal common ancestor; TI, thermophily index

* Tel.: +39-81-725-7313; fax: +39-81-593-6123.
E-mail address: digiulio@iigb.na.cnr.it (M. Di Giulio).

0378-1119/01/$ - see front matter
PII: S0378-1119(01)00781-8

program (Swofford, 1993). In order to reconstruct the ancestral sequences, the *states for interior nodes* option was used after building the specific topologies of the phylogenetic trees, selecting *accelerated transformation* (ACCTRAN) as the method for optimising characters (Swofford, 1993; Di Giulio, 2000b).

The reconstruction of the ancestral sequences by means of the maximum likelihood method was achieved using the ANCESTOR program of Zhang and Nei (1997). All the authors' recommendations were followed in using this program.

The thermophily index (TI) that can be associated to any one protein sequence has already been defined (Di Giulio, 2000a). Briefly, it is defined by the expression:

$$\mathrm{TI} = \sum_{j=1}^{N} R_j/N$$

where R_j is the value of the jth amino acid's thermophily rank (Di Giulio, 2000a), and N is the total number of amino acids in the considered protein (Di Giulio, 2000a).

The optimal growth temperature (T_{opt}) values of the various organisms were taken from Jacobs and Gerstein (1960) and from Staley et al. (1984). In certain cases, especially for eukaryotes, these values were found by consulting the specialised literature.

3. Results

Fig. 1 shows the correlation between the optimal growth temperature of the various organisms and the thermophily index (TI) (Di Giulio, 2000a) for 84 amino acid sequences: 47 from the signal recognition particle 54 kDa (SRP54) and 37 from the SR subunit α. The regression line ($T_{\mathrm{opt}} = -382.035 + 41.381\mathrm{TI}$) was highly significant ($F = 41.14$, df $= 83$, $P < 10^{-4}$).

It is clear that if we estimated the TI value for the ancestral sequence of the last universal common ancestor (LUCA) from and on this sequence set (Fig. 1), then the assignment of this value to mesophilic or (hyper)thermophilic sequences would be difficult because the mesophilic sequences extend over a large range of the thermophily index (Fig. 1). Therefore, I have removed 29 points from the correlation in Fig. 1 and obtained the result shown in Fig. 2, in which there is a clear separation between the mesophilic and (hyper)thermophilic sequences. Therefore, the assignment of the TI value of the LUCA's ancestral sequence (reconstructed from this sequence set (Figs. 2 and 3)) to the mesophiles or the (hyper)thermophiles should be less ambiguous in this case. This points removal can be justified by the fact that it does not in any way affect the estimate of the LUCA's ancestral sequence, provided that the set of sequences thus obtained is representative of the spectrum of the three main lines of divergence, and in this case it is. Furthermore, using the variability present in the original data, for instance in Fig. 1, to calculate the interval of confidence (see legend to Tables 1 and 2) makes this removal virtually inoffensive.

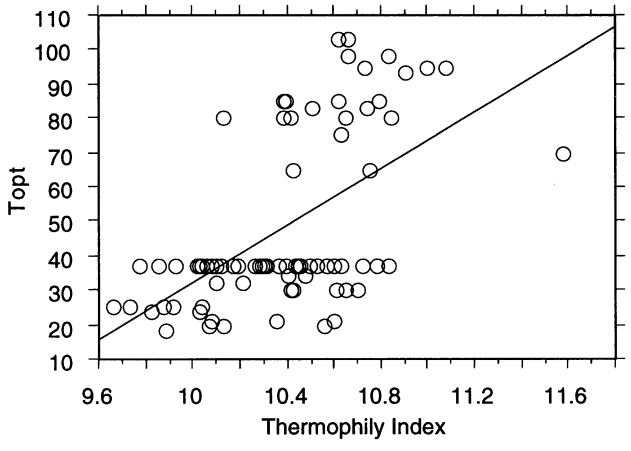

Fig. 1. Highly significant correlation between the optimal growth temperatures of the various organisms and the thermophily index for a total of 84 sequences of the two paralogous proteins (54 kDa and subunit α) of the signal recognition particle. See text for further information.

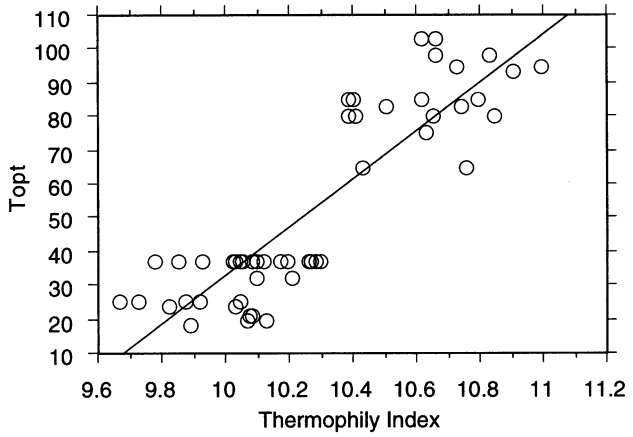

Fig. 2. Correlation obtained from Fig. 1 after removing 29 points from the correlation. See text for further information.

In theory, there should be no need to use the sequences of two paralogous proteins to reconstruct the LUCA's ancestral sequence but the sequences from a single protein might be sufficient. However, in practice this is not the case because the ANCESTOR program (Zhang and Nei, 1997) cannot work on rooted phylogenetic tree topologies and, thus, cannot estimate the sequence of the LUCA's node. The use of paralogous proteins removes this limitation and therefore makes it possible to estimate the LUCA's ancestral sequence, because the unrooted tree of the sequences of two paralogous proteins obviously contains two nodes for the LUCA corresponding to the deepest nodes of the set of sequences for every single orthologous protein.

The phylogenetic tree topologies that I have built are derived from that of the small subunit of ribosomal RNA (Maidak et al., 1997). For the set of sequences in Fig. 2, an example of these topologies is reported in Fig. 3 which

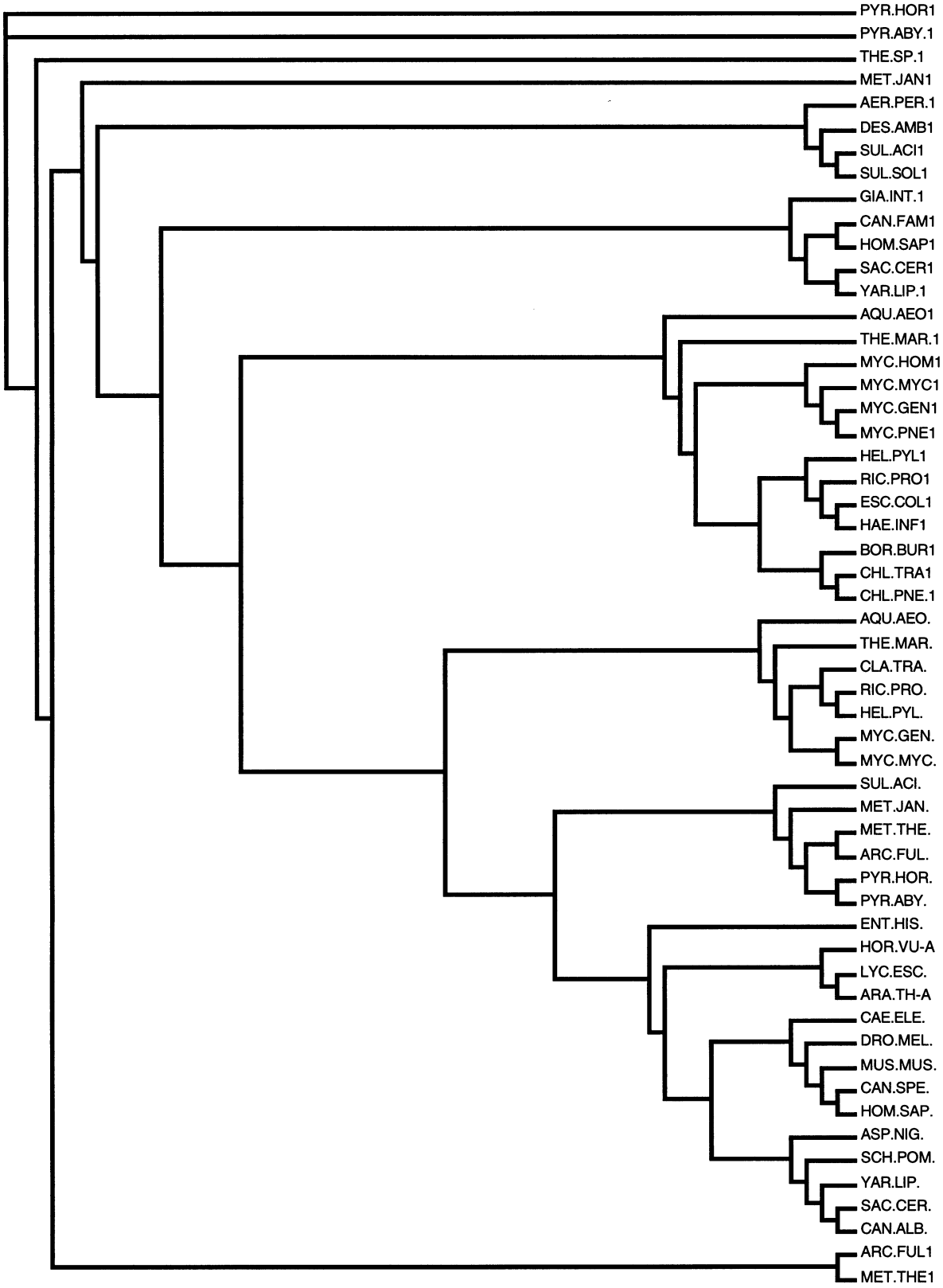

Fig. 3. The topology of one of the unrooted phylogenetic trees used in the analysis. The rooted topology for every single orthologous protein in the Bacteria domain is that of the ribosomal RNA of the small subunit (Maidak et al., 1997). To obtain the complete name of the species, see the web site mentioned in Section 2. The number '1' at the end of the organisms' names identifies the paralogous proteins in the α subunit, while the organisms with no number are those of the SRP54. See text for further information.

Table 1
Results of the analysis of the various ancestral sequences reconstructed using maximum parsimony and maximum likelihood for the paralogous proteins of the signal recognition particle: SRP54 and subunit α[a]

Ancestor	Maximum parsimony		Maximum likelihood		Paralogous proteins
	TI	T_{opt} (°C)	TI	T_{opt} (°C)	
Universal	11.004	104.4	10.698	82.6	SRP54
	11.068	109.0	10.893	96.5	SRα
Archaea/Eukarya	10.757	86.8	10.658	79.8	SRP54
	10.950	100.6	10.876	95.3	SRα
Bacteria	11.026	106.0	10.660	79.9	5RP54
	11.108	111.8	10.795	89.6	SRα
Archaea	10.723	84.4	10.784	88.8	SRP54
	11.114	112.3	10.873	95.1	SRα
Eukarya	10.194	46.8	10.394	61.0	5RP54
	10.410	62.1	10.236	49.7	SRα
Universal	10.718	84.1	10.739	85.6	SRP54
	10.942	100.0	10.843	93.0	SRα
Archaea/Bacteria	10.803	90.1	10.739	85.6	SRP54
	11.130	113.4	10.843	93.0	SRα
Bacteria	11.006	104.6	10.726	84.6	5RP54
	11.072	109.3	10.818	91.2	SRα
Archaea	10.845	93.1	10.742	85.8	SRP54
	11.193	117.9	10.831	92.1	SRα
Eukarya	10.232	49.5	10.380	60.0	SRP54
	10.253	51.0	10.091	39.4	SRα
Universal	10.911	97.8	10.586	74.7	SRP54
	10.980	102.7	10.853	93.7	SRα
Bacteria/Eukarya	10.753	86.6	10.586	74.7	SRP54
-	11.014	105.2	10.853	93.7	SRα
Bacteria	10.878	95.5	10.719	84.2	SRP54
	10.956	101.0	10.818	91.2	SRα
Archaea	10.833	92.3	10.715	83.9	SRP54
	11.273	123.6	10.831	92.1	SRα
Eukarya	10.227	49.1	10.307	54.8	SRP54
	10.383	60.2	10.120	41.5	SRα

[a] The possible topologies of the tree of life can be identified by the node of the ancestor with two domains, for example, the Archaea/Eukarya ancestor identifies the topology rooted in the Bacteria domain. The thermophily index (TI) is the value associated to the ancestral sequence of the indicated ancestor, whereas the optimal growth temperature (T_{opt}) is the one estimated from the corresponding TI value by means of the regression equation ($T_{opt} = -679.375 + 71.231$TI) in Fig. 2. The 95% interval of confidence (Wonnacott and Wonnacott, 1982) for these temperatures ranges from ±41.0 °C for the lowest TI value to ±42.5 °C for the highest. These intervals was calculated using data from Fig. 1. See text for further information.

shows an unrooted tree in which the Eukarya domain is the Archaea's sister domain. Moreover, the ancestral sequences were estimated using all three possible rootings of the tree of life that can be obtained from paralogous protein pairs.

Table 1 reports most of the information obtained from this analysis, i.e. the TI value for the reconstructed ancestral sequences of the ancestors we are interested in, and the estimate of the optimal growth temperature value that can be associated to these sequences (Table 1 and its legend).

I have carried out an equivalent analysis for the pair of paralogous proteins of the tryptophanyl- and tyrosyl-tRNA synthetases using a total of 49 sequences (Diaz-Lazcoz et al., 1998). Likewise for these sequences, a strong correlation is obtained ($F = 23.38$, df $= 48$, $P < 10^{-4}$) between the optimal growth temperature and the thermophily index (data not shown). The removal of 8 points from this correlation produces the regression shown in Fig. 4, whose points refer to 16 sequences of tryptophanyl-tRNA synthetase and

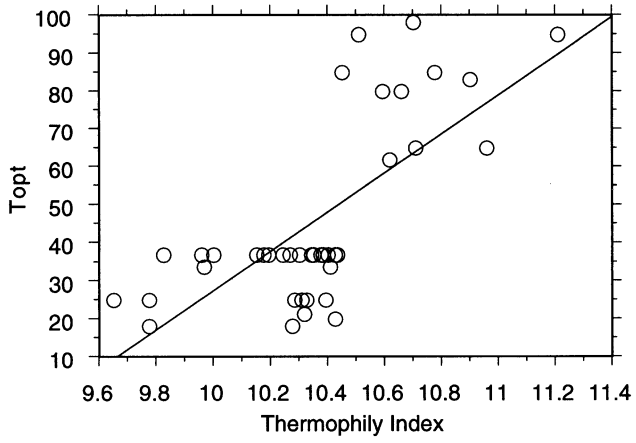

Fig. 4. Correlation between the optimal growth temperatures of the various organisms and the thermophily index for the pair of paralogous proteins of the tryptophanyl- and tyrosyl-tRNA synthetases. See text for further information.

Table 2
Results of the analysis of the various ancestral sequences reconstructed for the pair of paralogous proteins of the tryptophanyl- and tyrosyl-tRNA synthetases (TrpRS and TyrRS)[a]

Ancestor	Maximum parsimony		Maximum likelihood		Paralogous proteins
	TI	T_{opt} (°C)	TI	T_{opt} (°C)	
Universal	10.974	77.5	10.764	66.6	TrpRS
	11.051	81.5	10.776	67.2	TyrRS
Archaea/Eukarya	10.811	69.1	11.071	82.5	TrpRS
	11.006	79.2	10.858	71.5	TyrRS
Bacteria	10.480	51.9	10.488	52.4	TrpRS
	11.126	85.4	10.462	51.0	TyrRS
Archaea	10.959	76.7	10.870	72.1	TrpRS
	10.685	62.5	10.797	68.3	TyrRS
Eukarya	10.644	60.4	10.559	56.0	TrpRS
	10.898	73.6	10.929	75.2	TyrRS
Universal	11.049	81.4	10.724	64.6	TrpRS
	11.035	80.7	10.758	66.3	TyrRS
Archaea/Bacteria	10.961	76.8	10.809	69.0	TrpRS
	11.063	82.1	10.758	66.3	TyrRS
Bacteria	11.094	83.7	10.722	64.5	TrpRS
	10.466	51.2	10.522	54.1	TyrRS
Archaea	10.705	63.6	10.675	62.0	TrpRS
	11.114	84.8	11.098	83.9	TyrRS
Eukarya	10.724	64.6	10.620	59.2	TrpRS
	10.803	68.6	10.596	57.9	TyrRS
Universal	11.065	82.2	10.774	67.2	TrpRS
	10.917	74.6	10.840	70.6	TyrRS
Bacteria/Eukarya	11.146	86.4	10.774	67.2	TrpRS
	10.943	75.9	10.840	70.6	TyrRS
Bacteria	11.234	91.0	10.657	61.1	TrpRS
	10.526	54.3	10.423	49.0	TyrRS
Archaea	10.709	63.8	10.778	67.4	TrpRS
	11.059	81.9	11.120	85.1	TyrRS
Eukarya	10.819	69.5	10.567	56.4	TrpRS
	10.793	68.1	10.551	55.6	TyrRS

[a] For the meaning of the abbreviations, see the legend to Table 1. The regression equation that transforms the thermophily index (TI) value into the corresponding optimal growth temperature is $T_{opt} = -490.465 + 51.756TI$ (Fig. 4). The 95% interval of confidence (Wonnacott and Wonnacott, 1982) for these temperatures ranges from ±38.3 °C for the lowest TI value to ±40.6 °C for the highest. These intervals were calculated using data from the initial correlation of 49 sequences. See text for further information.

to 25 sequences of tyrosyl-tRNA synthetase. Finally, the building of three phylogenetic trees with all the possible rootings of the tree of life made it possible to reconstruct the ancestral sequences (from the set of sequences in Fig. 4) by means of both maximum parsimony and maximum likelihood and produced the temperature estimates reported in Table 2.

4. Discussion

The analysis of the pairs of paralogous proteins of the signal recognition particle clearly shows that the last universal common ancestor (LUCA) was a hyperthermophile 'organism' (Table 1). This is particularly true for the ancestral sequences estimated using maximum parsimony rather than for those derived by means of maximum likelihood (Table 1). Furthermore, this result is independent of how the tree of life is rooted. All three possible rootings give rise to a 'hot' LUCA (Table 1). Indeed, the set of sequences used in the present paper (Figs. 2 and 3) make it difficult to recover the alternative hypothesis of a mesophile LUCA. This can only be obtained when, by preserving the identity of the three domains, the sequences are ordered in such a way that the thermophily index (TI) values of the sequences in the phylogenetic tree topology (rooted in the Eukarya domain) range from the lowest value, close to the LUCA node, to the highest value, towards the less deep nodes (data not shown).

The ancestors of the Bacteria and Archaea domains are also hyperthermophiles (Table 1) while the ancestor of the Eukarya domain seems to be a mesophile (Table 1), above all if the TI values and hence the corresponding temperatures are seen on Fig. 2 and not estimated by means of the regression line, which raises these temperature values.

Overall, these observations are consistent with a large quantity of data, suggestions and theories (Woese, 1987; Achenbach-Richter et al., 1987; Wachtershauser, 1988,

1998; Pace, 1991; Holm, 1992; Stetter, 1995; Wiegel and Adams, 1998; Bocchetta et al., 2000; Nisbet and Sleep, 2001). However, the observation of a hyperthermophile LUCA clashes with the findings of an equivalent work (Galtier et al., 1999; see also Galtier, 2001) which found the LUCA to be a mesophile, although the data referred here agree with analyses of a similar nature (Di Giulio, 2000a,b). It is clear that only by perfecting the methods for reconstructing ancestral sequences, i.e. by better understanding molecular evolution, will we be able to give a 'definitive' answer to this question.

The observation of a non-mesophile LUCA is also confirmed for the pair of paralogous proteins of tryptophanyl- and tyrosyl-tRNA synthetases (Table 2), although in this case the LUCA seems to be more a thermophile than a hyperthermophile, at least for the estimates deriving from maximum likelihood (Table 2). (The anomalous behaviour of the Bacteria domain ancestor, which seems to be a mesophile (Table 2), can be explained by the fact that two sequences from hyperthermophiles (and in particular the sequence from *Aquifex aeolicus*) have a very low TI value (Fig. 4).)

In order to transform the LUCA from a (hyper)thermophile to a mesophile, it is sufficient in this case to transfer the sequence from *Aquifex aeolicus* from the position near the LUCA's node to positions close to the less deep nodes in the topology that sees the Bacteria domain as containing the root of the tree of life (data not shown). This occurs both for the estimates derived by maximum parsimony and those derived by maximum likelihood. In this case, therefore, the presence of hyperthermophile species close to the LUCA's node is fundamental to determining whether or not it is a (hyper)thermophile.

The average accuracy (Zhang and Nei, 1997) of the LUCA's ancestral sequences reconstructed using maximum likelihood (ML) has a mean value of 0.87 for the paralogous protein pair SRP54 and SR subunit α, and 0.76 for the tryptophanyl- and tyrosyl-tRNA synthetases. Therefore, the accuracy of the reconstruction for these deep nodes is relatively low. Furthermore, the comparison between the LUCA's ancestral sequences reconstructed with maximum parsimony (MP) and those reconstructed using ML indicates that they are relatively dissimilar, with a mean difference in the identity percentage of 21% for those of the SRP54 and SR subunit α, and 30% for those of tryptophanyl- and tyrosyl-tRNA synthetases. This seems to indicate that the two methods grasp different aspects of molecular evolution. An example of this is reported in Fig. 5, which compares two of these sequences. These refer to the LUCA's ancestral sequence, estimated using MP and ML, relative to the SR subunit α. The identity percentage between these sequences (Fig. 5) is 77.7% although the similarity percentage is 95.4% (Fig. 5). Overall, we can say that even if the sequences estimated by the two methods are relatively dissimilar, the conclusion that the LUCA was a thermophile or a hyperthermophile 'organism' is on the whole strengthened. This is because the difference between the ancestral sequences reconstructed by the two different methods seems to regard many aspects that should have been present in the true ancestral sequences.

Finally, whether or not the observation of a thermophile or hyperthermophile LUCA has any relevance in establishing that the origin of life took place at a high temperature is a question that has been comprehensively discussed elsewhere (Woese, 1987; Achenbach-Richter et al., 1987; Wachtershauser, 1988, 1998; Pace, 1991; Wiegel and Adams, 1998; Arrhenius et al., 1999; Galtier et al., 1999; Vogel, 1999; Di Giulio, 2000b; Nisbet and Sleep, 2001).

```
LUCA-MP   EKVIEFLEELEIALADVGVKVVVELIERLKKDELIIKILKEELRKILVIMFVGVNGSGKT
LUCA-ML   EKVVEVLEELRIALADVGVEVVEKIIEKIKEEETIKEALKEELLEILVIMFVGVNGSGKT
          ***:*.****.********:** ::**::*::* * : ***** :***************

LUCA-MP   TTIGKLAHYLKKKGKKVLLVAADTFRAAAVELEVLAERLAKEAVKKAKEEDVVIVDTAGR
LUCA-ML   TTIAKLAHYLKKKGKKVLLAASDTFRAAAIELEVLAERVAYDAIQHAKSRDVVIIDTAGR
          ***.***************.*:*******:********:* :*:::**..****:*****

LUCA-MP   LHVDKPLMEELQKIKKVIKPDEVLLVIDAMTGQDALEQAKAFKEAVITGIIVTKLDGSAK
LUCA-ML   LHTNKNLMEELKKINRVIKPDEILLVIDATTGQDAVNQAKAFNEAVITGIILTKLDGDAK
          **.:* *****:**::******:****** ******::*****:********:*****.**

LUCA-MP   GGAALSICYALGKPIKFIGVGEKIDDLRPFDPDAFVNRLL
LUCA-ML   GGAALSIAYALKKPIKFIGTGEKIDDLRPFDPERFVERLL
          *******.*** *******.************: **:***
```

Fig. 5. The ancestral sequences of the LUCA for the SR's subunit α, estimated using maximum parsimony (MP) and maximum likelihood (ML). These sequences respectively refer to the LUCA's nodes in the tree topology rooted in the Bacteria domain (Fig. 3). The star indicates that the same amino acid is present in the two sequences, the colon and the period respectively indicate similar or partially similar amino acids, while the white space indicates dissimilar amino acids.

Acknowledgements

Part of this work was carried out at the Institute for Theoretical Physics of the University of California at Santa Barbara and was supported by the National Science Foundation under Grant no. PHY 99-07949.

References

Achenbach-Richter, L., Gupta, R., Stetter, K.O., Woese, C.R., 1987. Were the original eubacteria thermophiles? Syst. Appl. Microbiol. 9, 34–39.

Arrhenius, G., Bada, J.L., Joyce, G.F., Lazcano, A., Miller, S., Orgel, L.E., 1999. Origin and ancestor: separate environments. Science 283, 792.

Bocchetta, M., Gribaldo, S., Sanagelantoni, A., Cammarano, P., 2000. Phylogenetic depth of the bacterial genera Aquifex and Termotoga inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. J. Mol. Evol. 50, 366–380.

Diaz-Lazcoz, Y., Aude, J.-C., Nitschké, P., Chiapello, H., Landés-Devauchelle, C., Risler, J.-L., 1998. Evolution of genes, evolution of species: the case of aminoacyl-tRNA synthetases. Mol. Biol. Evol. 15, 1548–1561.

Di Giulio, M., 2000a. The late stage of genetic code structuring took place at a high temperature. Gene 261, 189–195.

Di Giulio, M., 2000b. The universal ancestor lived in a thermophilic or hyperthermophilic environment. J. Theor. Biol. 203, 203–213.

Galtier, N., 2001. Maximum-likelihood phylogeny analysis under a covarion-like model. Mol. Biol. Evol. 18, 866–873.

Galtier, N., Lobry, J.R., 1997. Relationships between genomic G + C content, RNA secondary structures, and optimal growth temperature in prokaryotes. J. Mol. Evol. 44, 632–636.

Galtier, N., Tourasse, N., Gouy, M., 1999. A nonhyperthermophilic common ancestor to extant life forms. Science 283, 220–221.

Gribaldo, S., Cammarano, P., 1998. The root of the universal tree of life inferred from anciently duplicated genes encoding components of protein-targeting machinery. J. Mol. Evol. 47, 508–516.

Holm, N.G., 1992. Marine hydrothermal systems and the origin of life. Origins Life Evol. Biosph. 22, 1–241.

Jacobs, M.B., Gerstein, M.J., 1960. Handbook of Microbiology, van Nostrand, London.

Maidak, J.L., Olsen, G.J., Larsen, N., Overbeek, R., McCaughey, M.J., Woese, C.R., 1997. The RDP (ribosomal database project). Nucleic Acids Res. 25, 109–110.

Nisbet, E.G., Sleep, N.H., 2001. The habitat and nature of early life. Nature 409, 1083–1091.

Pace, N.R., 1991. Origin of life—Facing up to the physical setting. Cell 65, 531–533.

Staley, J.T., Bryant, M.P., Plennig, N., Holt, J.G., 1984. In: Hensyl, W.R. (Ed.). Bergey's Manual of Systematic Bacteriology, Vol. 3. Lippincott Williams and Wilkins, Philadelphia, PA.

Stetter, K.O., 1995. Microbial life in hyperthermal environments. ASM News 61, 285–290.

Swofford, D.L., 1993. PAUP: Phylogenetic Analysis Using Parsimony, version 3.1.1, Laboratory of Molecular Systematics, Smithsonian Institution, Washington, DC.

Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F., Higgins, D.G., 1997. The CLUSTALX windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. Nucleic Acids Res. 25, 4876–4882.

Wachtershauser, G., 1988. Before enzymes and templates: theory of surface metabolism. Microbiol. Rev. 52, 452–484.

Wachtershauser, G., 1998. The case for a hyperthermophilic, chemolithoautotrophic origin of life in an iron-sulfur world. In: Wiegel, J., Adams, M.W.W. (Eds.). Thermophiles: The Keys to Molecular Evolution and the Origin of Life? Taylor and Francis, London, pp. 47–57.

Wiegel, J., Adams, M.W.W. (Eds.), 1998. Thermophiles: The Keys to Molecular Evolution and the Origin of Life? Taylor and Francis, London.

Woese, C.R., 1987. Bacterial evolution. Microbiol. Rev. 51, 221–271.

Wonnacott, T.H., Wonnacott, R.J., 1982. Introductory Statistics, Wiley, New York, pp. 281–304.

Vogel, G., 1999. RNA Study suggests cool cradle of life. Science 283, 155–157.

Zhang, J., Nei, M., 1997. Accuracies of ancestral amino acid sequences inferred by the parsimony, likelihood, and distance methods. J. Mol. Evol. 44 (Suppl. 1), S139–S146.